Concentration of measure (tail bounds)

Today
- quick recap of tail bounds
- application to balls in bins
- Poisson approx for analyzing balls in bins
- estimating reachability

The more you know about r.v. X, the better the bounds

① If only thing you know is mean & $X \geq 0$ (but not $= 0$)

Markov's Inequality

$$\Pr(X \geq t) \leq \frac{E(X)}{t} \qquad t > 0 \qquad \left[\Pr(X \geq c\,E(X)) \leq \frac{1}{c}\right] \quad c > 0$$

Corollary

If $0 \leq X \leq B \implies \Pr(X \leq t) \leq \frac{E(B-X)}{B-t}$

② If you know mean μ and variance σ^2

Chebychev's Inequality

$$\forall t > 0 \qquad \Pr(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2}$$

σ: std deviation

③ Sums of indep r.v.'s

Theorem $X = X_1 + X_2 + \dots + X_n$ where X_i's indep & $X_i \in [0,1]$ with $E(X_i) = p_i$

$$\mu = E(X) = \sum_{i=1}^{n} p_i \qquad p = \frac{\mu}{n}$$

$$\boxed{\Pr(X \geq \mu + \lambda) \leq e^{-n H(p + \frac{\lambda}{n} \| p)}} \quad (1)$$

$$\boxed{\Pr(X \leq \mu - \lambda) \leq e^{-n H(1-p+\frac{\lambda}{n} \| 1-p)}} \quad (2)$$

$$\boxed{H(x \| p) = x \ln\left(\frac{x}{p}\right) + (1-x)\ln\left(\frac{1-x}{1-p}\right) \quad \text{relative entropy}}$$

More useful forms (corollaries)

$$\boxed{\begin{array}{l} \Pr(X \geq \mu+\lambda) \\ \Pr(X \leq \mu-\lambda) \end{array} \leq e^{-\frac{2\lambda^2}{n}}} \quad (3)$$

$$\boxed{\begin{array}{ll} \Pr\left(X \leq (1-\delta)\mu\right) \leq e^{-\frac{\delta^2 \mu}{2}} & 0 \leq \delta \leq 1 \\ \Pr\left(X \geq (1+\delta)\mu\right) \leq e^{-\mu\left[(1+\delta)\ln(1+\delta)-\delta\right]} \leq \begin{cases} e^{-\frac{\delta^2\mu}{2+\delta}} & \delta > 0 \\ e^{-\frac{\delta^2\mu}{3}} & 0 < \delta \leq 1 \end{cases} & \end{array}} \quad (4)$$

$\mu_L \leq \mu$

$\mu \leq \mu_H$

$$\boxed{\begin{array}{l} X = X_1 + X_2 + \dots + X_n \\ X_i\text{'s indep.} \quad X_i \text{ takes values in } [a_i, b_i] \quad E(X) = \mu \\ \begin{array}{l} \Pr(X \leq \mu-\lambda) \\ \Pr(X \geq \mu+\lambda) \end{array} \leq e^{-\frac{2\lambda^2}{\sum_{i=1}^{n}(b_i-a_i)^2}} \end{array}} \quad (5)$$

Comments:

- many different versions (hold for Poisson r.v.s, Gaussian etc); we'll see more later.
- also hold for "negatively correlated r.v.s" [negatively associated]

$$E\left[e^{t(X+Y)}\right] \leq E\left[e^{tX}\right] E\left[e^{tY}\right]$$

Example

- edges in random spanning trees

One more simple probabilistic tool

Union Bound

Let $E_1, E_2, \dots, E_k$ be any collection of events in prob space

Then $\Pr(E_1 \vee E_2 \vee \dots \vee E_k) \le \sum_{i=1}^{k} \Pr(E_i)$

Application: Balls in Bins

n bins

Throw n balls, one at a time, independently at random into a bin. Prob $\to i^{th}$ bin $= \frac{1}{n}$

$B_i = \#$ balls in i^{th} bin at end

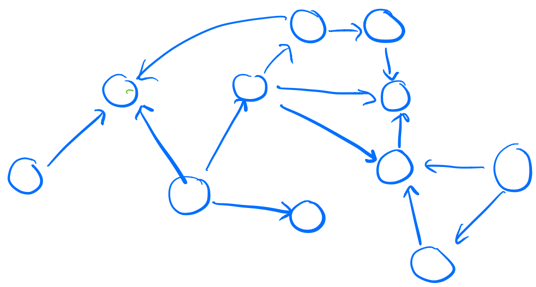
Theorem $\Pr\left(\max_i B_i \geq e \frac{\ln n}{\ln \ln n}\right) \le \frac{1}{n^c}$ for some $c > 0$

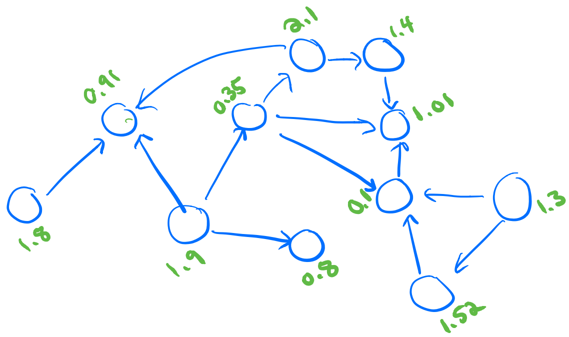
proof idea: prove that prob that a particular bin so unlikely to exceed this bound that a union bound gives result

A cool application: reachability

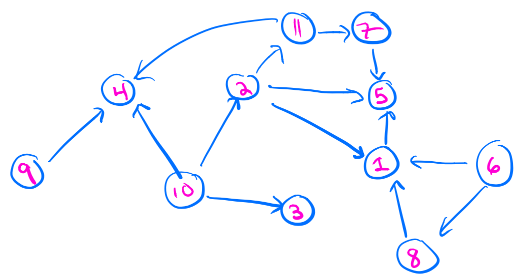
Input: directed graph $G = (V, E)$

Output: $\forall v \in V$, $|S(v)|$ where $S(v)$: set of nodes reachable from v

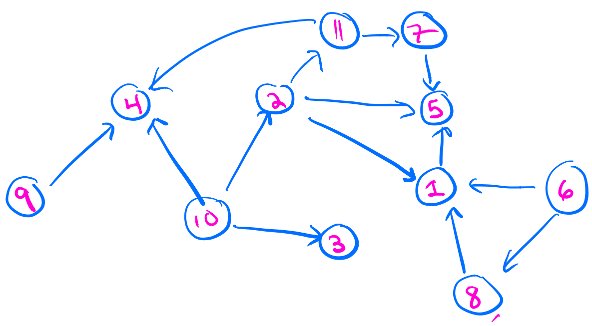




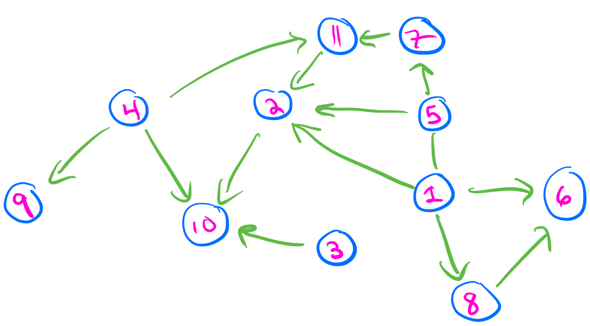
2.1
1.4
0.91
0.35
1.01
1.3
1.8
1.9
0.8
1.52



11
7
4
2
5
9
10
1
6
3
8



11
7
4
2
5
9
10
1
6
3
8



11
7
4
2
5
9
10
1
6
3
8